U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

DOT HS 807 796
NHTSA Technical Report

December 1991

National Accident Sampling System General Estimates System Technical Note, 1988 to 1990

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT-HS-807-796		

4. Title and Subtitle	
National Accident Sampling System General Estimates System Technical Note, 1988 to 1990	5. Report Date December 1991 6. Performing Organization Code NRD-31 8. Performing Organization Report No.
7. Author(s) Terry S. T. Shelton	
9. Performing Organization Name and Address Mathematical Analysis Division National Center for Statistics and Analysis 400 7th St. S.W. Washington, D.C. 20590	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Research and Development National Highway Traffic Safety Administration 400 7th St. S.W. Washington, D.C. 20590	13. Type of Report and Period Covered NHTSA Technical Report 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

The National Accident Sampling System General Estimates System (GES) is operated by the National Center for Statistics and Analysis, an office within the National Highway Traffic Safety Administration's Research and Development. GES is a probability sample of approximately 45,000 motor vehicle police traffic accident reports selected on an annual basis. This Technical Note explains how estimates are derived from GES data and how reliable those estimates are. The set of crashes described by GES estimates, the sample selection procedures, the estimation procedure, and the reliability of the estimates in terms of sampling and nonsampling error are discussed.

17. Key Words	18. Distribution Statement
sample, stratum, sampling error, nonresponse	Document is available to the public through the National Technical Information Service Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

APPENDICES

I. INTRODUCTION

The National Highway Traffic Safety Administration (NHTSA) uses data from many sources to support its motor vehicle highway safety research and development programs. One of its national data sources is the National Accident Sampling System General Estimates System (GES). GES is operated by the National Center for Statistics and Analysis (NCSA), an office within NHTSA's Research and Development. GES is a survey of police-reported traffic crashes in the country.

Estimates from GES data are used to identify highway safety problem areas, provide a basis for regulatory and consumer information initiatives, and form the basis for cost and benefit analyses of highway safety initiatives.

Within the constraints of police-reported data, GES estimates are intended to answer general questions on a national level about motor vehicle crashes:

When and how often do crashes occur?
Where do crashes occur?
What happened during the crash?
Was alcohol involved?

the vehicles involved:

What types of vehicles are involved in crashes?
What is their action prior to the crash?
What area of the vehicle is damaged in the crash?

and the people involved:

> How many drivers, passengers, pedestrians and pedalcyclists were involved?
> How severely were they injured?
> What are their age and sex?
> Were they wearing seat belts?

In addition, GES was intended to provide a means to track trends in these national level estimates.

The purpose of this technical note is to explain how the estimates from GES data are derived and how reliable they are. The set of crashes described by GES estimates, the sample selection procedures, the estimation procedure and the reliability of these estimates are discussed. This paper covers the first three years of GES, 1988 to 1990, and will be updated each year. Questions about GES that have not been answered in this report, or requests for more information about GES, may be addressed to:

General Estimates System, NRD-31
National Highway Traffic Safety Administration
400 Seventh St., S.W.
Washington, D.C. 20590
(202) 366-5362

ABBREVIATIONS

Throughout this report, many abbreviations have been used to simplify reading. Each is defined when the abbreviation is first used. However, a list of abbreviations with brief explanations has been compiled for reference and can be found in Appendix A.

II. SAMPLING FRAME AND DEFINITIONS

The General Estimates System is based on a probability sample of approximately 45,000 motor vehicle police traffic accident reports selected on an annual basis. The crashes eligible for the GES sample, called the sampling frame, are all motor vehicle traffic crashes in the United States that meet the following criteria:

- The crash must involve a harmful event. This means that the PAR must show that personal injury or property damage resulted from the crash; and,

- The crash must involve at least one motor vehicle in transport. This means that at least one vehicle must have been in motion on a public trafficway or, if not in motion, at least one vehicle must have been partially on or over the roadway.

- A police accident report (PAR) must be filled out by a police officer and eligible to be sent to the state's agency responsible for recording crashes;

Crashes of all severities-- from property damage only to fatal-- are included in the GES sampling frame. All motor vehicles-- automobiles, automobile derivatives, light trucks and vans, utility vehicles, medium and heavy trucks, motored cycles and all-terrain vehicles-- are included in the GES sample. The complete list of vehicles is shown in Appendix B. The total number of crashes within scope of GES, as estimated from GES data, was approximately 6.5 million in 1990.

Other sources, such as the National Safety Council, estimate that approximately 12.8 million motor vehicle traffic crashes occurred in

1990. This number includes those crashes not reported to the police or by the police and usually involve only minor property damage and no significant personal injury. By restricting attention to police-reported crashes, GES concentrates on those crashes of greatest concern to the highway safety community and the general public. But more importantly, police-reported crashes can be easily identified and measured, unlike unreported crashes.

Although police-reported crashes can be identified, the definition of a police-reported crash in one state may be very different than another state's definition. Some states require crashes to be reported by police based on a minimum amount of property damage, such as $500 in Texas. Other states base their reporting threshold on whether or not any of the involved vehicles had to be towed or whether or not any involved persons were injured, such as Pennsylvania.

A more detailed discussion of definitions used to determine crash qualification criterion for GES can be found in the "1990 Crashworthiness Data System Coding and Editing Manual." [1/]

III. SAMPLE DESIGN

The selection of PARs for GES is accomplished in three stages.

FIRST STAGE

The first stage is a sample of geographic areas from across the United States called Primary Sampling Units (PSUs). The PSUs were formed originally by the Highway Safety Research Institute (now the University of Michigan Transportation Research Institute) and were based on 1973 population estimates. A PSU can be a central city, a county surrounding a central city, an entire county, or a group of contiguous counties.

Before the PSUs were selected they were arranged into different geographic and urbanization classes, called strata, to be sure that each class is sufficiently represented in the sample. Stratification can also reduce the errors in the estimates that occur due to sampling. Different stratification choices were investigated, but the following 12 strata were chosen because estimates of sampling errors were lowest for these strata. The PSUs were assigned to these strata according to PSU geographic region and PSU type:

- Geographic Region - Northeast, South, Central, and West.

- Type - Large Central City, Large Suburban Area, All Others.

Sixty PSUs were selected for the first stage GES sample with probabilities proportional to the number of fatal and injury crashes in each PSU. These numbers were obtained from 1983 state publications. Table 1 shows the number of selected PSUs and total PSUs in the country by stratum. The selection of the GES PSUs, however, can be considered a continuation of the original 1978 NASS PSU sample. To completely understand the GES PSU sample, therefore, an overview of the NASS PSU sample is necessary. For a thorough discussion of the original NASS sample design, refer to "National Accident Sampling System Sample Design, Phases 2 and 3." 2/

TABLE 1: GES PSU SAMPLE AND FRAME BY STRATUM

REGION SIZE	FRAME	SAMPLE
Northeast		
Central City	19	4
Suburban	69	8
Other	94	2
Midwest		
Central City	15	4
Suburban	79	8
Other	280	4
South		
Central City	21	4
Suburban	87	8
Other	368	6
West		
Central City	12	2
Suburban	42	8
Other	109	2
TOTAL	1,195	60

The NASS PSU sample was designed by Westat Inc., an NCSA contractor, and was intended to provide detailed crash data on a probability sample of crashes. The original design called for 75 PSUs to be phased in over several years, starting with a pilot study with 10 PSUs in 1978. All 75 PSUs were selected in 1977 from a total of 1279 PSUs using 1977 population as the measure of size. NASS began with 10 PSUs in 1978, ten more PSUs were added in 1980, ten more in 1981, and from 1982 through the first eight months of 1986, 50 PSUs were maintained.

In August 1986, an agency decision was made to redesign the NASS into two different systems that would begin operation in January, 1988: the NASS Crashworthiness Data System (CDS) and the NASS GES. The CDS PSU sample was partially selected from the "old" NASS sample of 50 PSUs. A probability-based procedure called Keyfitzing was used because it maximizes the overlap between successive samples from the same or similar population. [3/] Thirty of the 50 PSUs were maintained using this procedure. Prior to the mid-year reselection, new PSU measures of size were calculated using the police-reported fatal and injury crash totals from 1983. Each state's annual crash publication was reviewed and counts of crashes per county were used to create these new measures of size. During this process, some PSUs with only a small number of crashes were grouped together and the number of PSUs in the country was redefined to be 1,195.

A larger PSU sample size was required for GES than for CDS because GES estimates were to be used to detect the usually small year-to-year changes in national crash estimates. With the first three years of GES data, certain trends have been identified with these data. For example, the number of severe or fatal injury crashes has declined by eight percent and the number of crash-involved motorcycles has decreased by 14 percent from 1988 to 1990. In addition to precision requirements, cost constraints played a major role in determining the GES PSU sample size. A GES PSU sample size of 60 was compromise between cost and precision constraints. Appendix C is a list of the GES PSUs.

The GES PSU sample consisted of the 36 CDS PSUs plus 24 newly selected PSUs. The additional 24 PSUs were allocated so that each stratum would contain an even number of PSUs. The selection was based on probabilities proportional to the number of fatal and injury crashes in that PSU. The more crashes in a PSU, the greater the likelihood that PSU would be included in the sample. The first stage sampling weights (the inverse of the PSU probabilities of selection) are shown in Appendix D.

Following the selection of new PSUs in 1986, much of 1987 was spent establishing the new PSUs and developing the new collection methodology. In 1988, NASS began anew as the two different systems, NASS GES with 60 PSUs and CDS with 36 PSUs. In 1991, the CDS PSU sample was reduced to 24 PSUs. More information will be published on the CDS in a separate technical note.

SECOND STAGE

The second stage in the GES sample selection process is a sample of police jurisdictions (PJ) within the PSUs. In most PSUs, the number of PJs is more than can be reasonably visited by a data collector. All PJs within a PSU were enumerated and the number of crashes investigated by each was determined. A probability sample of PJs within each PSU was selected with probability proportional to the number of crashes investigated by that PJ. The more crashes investigated by the PJ, the greater the likelihood that jurisdiction would be chosen. An average of seven PJs were selected within each PSU. The sampling weights for each selected PJ in the 60 PSUs (the inverse of the PJ probabilities of selection) are shown in Appendix D.

The first two stages in the GES sample selection process were completed prior to the 1988 survey year. The sampled PSUs and PJs are not reselected each year because there is a very large cost and a long time needed to establish new PSUs and develop cooperation from new PJs in a PSU. The third and final stage of selection is the only stage where new sampling units are introduced each year and where year-to-year variation will occur.

THIRD STAGE

The final stage of sampling is the selection of the PARs within the sampled PJs. The GES data collectors make weekly, biweekly, or monthly visits to each of the PJs in the sample. During the visit, the data collectors identify all PARs not reviewed on previous visits. Each PAR is reviewed to determine what type of vehicles were involved, whether or not any of these vehicles were towed, and whether or not someone was injured. A list of PARs is made with each PAR being assigned to different groups or strata based on the review criteria. The PARs are listed in the order they are available to the collector, which is assumed to be random on the day of listing. The PARs are sampled at different rates based on the stratum they have been assigned. The PAR strata at this third stage of sampling should not be confused with the PSU strata that were used in the first stage of the GES sample.

In 1988 and 1989, the PARs were stratified as follows:

- Stratum 1: All crashes involving a passenger vehicle, (passenger car, light truck, light van, or utility vehicle), that was towed from the scene of the crash due to damage;

- Stratum 2: All crashes not involving a towed passenger vehicle, but one in which an involved person was injured; and,

- Stratum 3: All other crashes.

This stratification was designed so that listing of crashes for the GES could be done at the same time as CDS listing. As mentioned earlier, 36 of the 60 GES PSUs are also CDS PSUs. In those PSUs, the CDS and GES PJs are the same and the same personnel list and sample crashes for CDS and GES. To save time and money, the stratification of PARs was designed to serve GES and CDS sampling purposes.

In addition, this stratification allows different sampling rates to be established for each stratum. For example, more serious crashes are those in Strata 1 and 2. These crashes must involve a towed passenger vehicle or personal injury. Since these crashes are of generally of greater interest to the highway safety community, these were sampled at a higher rate (over sampled) than if simple random sampling of PARs was employed. On the other side, about more than two-thirds of eligible crashes for GES involve only possible injury or no injury--generally Stratum 3 crashes. A non-stratified sample of PARs would result in two-thirds of these lower severity crashes. By adjusting the rates for Stratum 3, though, only about one-third of the crashes sampled were from Stratum 3.

Stratification can also help to reduce sampling error by allowing oversampling of Stratum 1 crashes. The characteristics of the crashes in Stratum 1 can be quite different-- ranging from a passenger car hitting a guardrail with no passenger injuries to a utility vehicle rolling over, killing its driver. Due to this variation, Stratum 1 crashes were over sampled. More crashes were chosen from this stratum than would have been if the rates were based on number of crashes in each stratum alone.

Within each stratum, a systematic sample of crashes is selected, based on different sampling rates. The rates were calculated prior to the survey year making use of historical knowledge of the number of crashes falling into each stratum and the interests of the users. The sampling rates for each PAR within a Stratum were calculated to produce an overall probability of selection that is dependent only on the stratum and not on the PSU or PJ. Across time, these listing and sampling procedures result in a systematic sample within each PSU, PJ, and stratum.

Regardless of the PSU and PJ, the national rate of selection for 1988 and 1989 were approximately:

Stratum 1 - 1 in every 100;
Stratum 2 - 1 in every 30; and,
Stratum 3 - 1 in every 280.

The distribution of sampled PARs along with the estimated number of PARs in the country in each stratum are shown in Table 2:

TABLE 2: SAMPLE SIZE AND ESTIMATED FRAME SIZE BY STRATUM FOR 1988 AND 1989

	STRATUM 1	STRATUM 2	STRATUM 3	TOTAL
1988	22,361 (2,471,000)	6,355 (267,000)	19,712 (4,138,000)	48,698 (6,877,000)
1989	23,245 (2,414,000)	6,858 (264,000)	14,002 (3,967,000)	44,105 (6,645,000)

The national rate desired, the PSU weight (inverse of the probability of selecting the PSU), and the PJ weight (inverse of the probability of selecting the PJ) are used to obtain sampling intervals for data collectors to actually apply while listing PARs at the PJs:

$$\text{Interval}_s = 1/[(\text{National Rate})^*\text{Weight}_{psu} * \text{Weight}_{pj}],$$

where s = Stratum 1, 2 or 3,
$\text{Weight}_{psu} = 1/P_{psu\ selection}$, and
$\text{Weight}_{pj} = 1/P_{pj\ selection}$.

These intervals represent the third stage weight for the sampled PAR and are used to select the actual sample of PARs. For example, using the PSU and PJ weights given in Appendix C, the interval for Stratum 1 in PJ 1 in PSU 16 would be:

$$\text{Interval}_1 = 1/[(1/100)^* \ 1.88 * 10] = 5.32.$$

In this case, the interval would be rounded to 5. So for PSU 16, PJ 1, every fifth PAR listed as a Stratum 1 PAR would be sampled. Whole numbers are used to avoid errors in selecting the sample. With whole numbers, line numbers can be used to correspond with sampling intervals. When the line number equals the sampling interval, a PAR is selected. Then the line numbering starts again at one. Data collection is discussed more in the next section.

The product of the PSU weight and the PJ weight may be more than the inverse of the national rate in some cases. For example, there are so few stratum 2 crashes, the intervals calculated are usually less than one. An interval less than one would mean sampling more crashes than actually occur in that PJ for that stratum. Since that is impossible, if the intervals for any stratum are less than 1, they are automatically set to 1. The net result of this adjustment is that fewer crashes than desired are sampled.

In 1990 the sample strata were changed slightly to accommodate a larger sample of medium and heavy truck crashes requested by the Federal Highway Administration. These strata were:

- Stratum 1: All crashes involving a towed passenger vehicle, (passenger car, light truck, light van, or utility vehicle), that was towed from the scene of the crash due to damage but no medium or heavy trucks were involved;

- Stratum 2: All crashes involving a medium or heavy truck and where at least one passenger vehicle was towed or an involved person was injured;

- Stratum 3: All crashes not involving a towed passenger vehicle, or a medium or heavy truck, but one in which an involved person was injured; and,

- Stratum 4: All other crashes.

The national rates of selection for 1990 were approximately:

Stratum 1 - 1 in every 119;
Stratum 2 - 1 in every 16;
Stratum 3 - 1 in every 28; and,
Stratum 4 - 1 in every 285.

The distribution of sampled PARs along with the estimated number of PARs in the country in each stratum are shown in Table 3:

TABLE 3: SAMPLE SIZE AND ESTIMATED FRAME SIZE BY STRATUM FOR 1990

STRATUM 1	STRATUM 2	STRATUM 3	STRATUM 4	TOTAL
19,021 (2,240,000)	6,679 (104,000)	6,970 (233,000)	13,620 (3,885,000)	46,290 (6,462,000)

SUBSAMPLING

In some of the larger PJs, too many crashes occur to be reasonably listed by the data collector. In these PJs, the data collector lists only a subset of the PARs and samples from them. The subset of crashes is chosen randomly, usually based on the PAR number. For example, in one large jurisdiction only PARs ending with the number 1 are listed, introducing a subsample weight of 10. If subsampling occurs at the PJ, the PJ probability of selection includes the subsampling probability of selection. These "final" PJ weights are shown in Appendix D.

IV. DATA COLLECTION

The GES data collection process is much simpler than the sampling process. The PARs are stratified and listed in the appropriate columns of the GES Stratification Record (SR). (See Appendix E for the 1990 version.) The four columns on the SR correspond to the four strata described in the previous section about the third stage of sampling.

As a PAR is listed and its stratum determined, sequential line numbers are entered in the appropriate column. When the line number is the same as the assigned sampling interval, that PAR is highlighted as the sampled PAR. Once the sampled PARs are identified, copies of each are made and sent to a data coding contractor. No other data are collected. No on-scene investigations or passenger interviews are conducted. No hospital data, vehicle registration data, or driver license data are collected. Once the sampled PAR has been copied, the data collection process ends.

For more detailed information on data collection, refer to the "Crashworthiness Data Collecting, Coding, and Editing Manual" and the "GES Researcher Manual, 1/90." [4/]

V. DATA CODING AND COMPLETENESS

The sampled GES PARs are sent to a coding and quality control contractor where trained personnel interpret and code the GES data directly onto an electronic file. The data are obtained either directly from an item on the PAR or by interpreting the information provided on the report in the accident diagram, the officer's written summary of the crash, or combinations of variables.

Each of the 26 states and three of the cities in the GES sample have different PAR formats. For example, in some states "initial point of impact" has ten possible codes and in other states there are 16 possible codes. Coding the 80 GES variables from these differing formats can be a difficult task. The data entry system accommodates these differences by actually being 29 different data entry systems. Each system assists the data coder by indicating where, on that particular PAR, the data can be found and how each of the state's codes can be translated into GES codes. Coding manuals have been developed for each of the different PAR formats.

Data are entered in five separate data entry sets: crash, vehicle, driver, occupant, and nonmotorist. When the GES datafiles are created after coding is complete, the occupant and nonmotorist data entry sets are combined and the driver and vehicle data entry sets are combined. The GES datafile, therefore, is really three GES files: the crash file, the vehicle/driver file, and the person file. A complete list of the 1990 GES variables in each of these files is shown in Appendix F.

MISSING DATA

If the GES data cannot be obtained from the PAR, the coder enters the unknown code, usually a "9" or "99". Unknown data usually result from these situations:

- The police officer failed to fill out the information on the PAR or the information was not legible and the information could not be derived from the diagram, the narrative, or other variables; or,

- The PAR does not have a block to fill in the data and the information could not be derived from the diagram, the narrative, or other variables.

- The police officer enters checks the unknown box for a variable on the PAR.

Two types of GES variables can never be coded unknown: yes/no variables and computer-assigned variables. For yes/no variables-- such as *fire occurrence*, *jackknife occurrence*, and *rollover occurrence*-- it is assumed if no mention of fire, jackknife or rollover is made that these events did not occur. Computer-assigned variables such as *PSU*, *region* and *sampling weight*, never have unknowns.

For some variables, unknowns are rarely coded because the police officer seldom enters an unknown code on the PAR. For example, the variable *injury severity* was only coded unknown for five percent of the persons on the 1990 GES Person file. All PAR formats have this variable, so it is usually not coded unknown as a result of missing or illegible data. It is coded unknown when the police officer codes it unknown.

Appendix F shows the percentage of unknowns for each variable for 1990. Making estimates with unknown data is discussed in the next section.

VI. ESTIMATION METHODS

In order to calculate estimates of national crash characteristics, data from each sampled PAR must be weighted to reflect its probability of selection. Because there are three stages in the GES sampling process, the final sampling weight is the product of the inverse of the probability of selection at each of these stages. The sum of the final sampling weights for all sampled PARs with the characteristic of interest produces the estimate of the national total for that characteristic. The general formula of the estimator for a population total for crash level variables is:

$$Y = \sum_{i=1}^{n} Y_i / P_i$$

where n = PAR sample size.

Y_i = 1 if ith PAR has the characteristic of interest, and 0 if not.

P_i = the probability of including the ith PAR in the sample.

P_i is the product of the probabilities of selection at each stage of sampling:

$$P_{\text{psu selection}} * P_{\text{pj selection}} * P_{\text{par selection}}$$

For example, using the PSU and PJ weights from Appendix C for PSU 16, PJ 1, and strata 1:

$$P_{\text{psu selection}} = 1/10$$
$$P_{\text{pj selection}} = 1/1.88$$
$$P_{\text{par selection}} = 1/5$$

the total P_i for that PAR would be

$$P_i = 0.1 * 0.532 * 0.2 = 0.0106.$$

That PAR, if it had the characteristic of interest, would add 1/.0106 or 94 to the estimate of the total.

Similar formulas would be used for estimates of vehicle or person level characteristics. For vehicle estimates, each vehicle with the characteristic of interest in the crash would be weighted by P_i. Similarly, for person estimates, each person with the characteristic of interest in the crash would be weighted by P_i.

NONRESPONSE ADJUSTMENT

For most published tables, known GES data have been used to calculate percentages, and adjustments have been made to the totals to account for the unknown data. For example, suppose the number of drivers is estimated to be:

Male Drivers	6,500
Female Drivers	3,200
Drivers of Unknown Sex	300
Total	10,000

If the assumption is made that the sex of the unknown drivers is in same proportion as the known drivers, the estimate for male drivers would become:

$$6{,}500 + [6{,}500/(6500+3200-300)] * 300 = 6{,}701.$$

The distributions of characteristics for the unknowns may not always be similar to the knowns. For variables with a very high percentage of unknowns, making this assumption could lead to misleading estimates, if in fact the unknowns are very different from the knowns. For the 1988 [5/], 1989 [6/] and 1990 [7/] GES Reports, only four of the 38 variables used in the reports had unknowns of more than 10 percent.

More sophisticated nonresponse adjustment methods are being developed for GES. For example, a hot-decking [8/] procedure has been investigated for several of the variables critical to the NHTSA analyses. This procedure replaces unknown information for a crash with known information from a similar crash. For example, speed limit is missing for almost one-third of all crashes in the GES accident file. For a particular crash with missing speed limit ("bad" crash), the accident file would be searched for another crash ("good" crash) where a set of correlated variables were the same as the "bad" crash and were known. Speed limit from the "good" crash would replace the missing speed limit of the "bad" crash. Correlated variables for speed limit that could be used are number of travel lanes, roadway alignment, interstate highway, and roadway profile. Plans are to implement new imputation procedures for at least some variables in the 1991 GES.

After the implementation of imputation methods, the files will be made available to the public with both the original variables and the imputed variables. The analyst using the files has the option of making estimates with or without the imputed variables. This procedure allows the analysts to make their own decisions to use a variable with a large percentage of unknowns.

VII. RELIABILITY OF ESTIMATES

Estimates produced from GES data are derived from an annual probability sample of approximately 45,000 PARs, not from a census of all 6.5 police-reported million crashes in the U.S. Consequently, the estimates are subject to sampling errors, as well as nonsampling errors.

SAMPLING ERRORS

Sampling errors are the differences that can arise between results derived from a sample and those computed from observations of all units in the population being studied. Since GES data are derived from a probability sample, estimates of the sampling error can be made. However, due to the complex sample design of GES, no textbook formulas exist for calculating sampling errors directly. Instead, estimates of GES sampling errors were calculated using a statistical computer program, WESVAR, [9/] employing the method of balanced repeated replicates (BRR). [10/] For every estimate in the 1988 GES Report, an estimate of the sampling error was calculated.

Publishing sampling error estimates for every GES estimate would be prohibitive in publication and cost constraints because the GES publication would double in size. A more common approach to displaying sampling error data for large publications that is used by the U.S. Census Bureau and the Bureau of Labor Statistics is to create a generalized variance model. The desirable characteristics of a generalized variance model are:

- The published data should be easy to use; and,
- The model should be reasonably accurate in estimating sampling errors.

The sampling error estimates generated by WESVAR for GES estimates were tested in several different models using regression techniques. In each model tested, the estimate of variance (sampling error squared) was the dependent variable and the estimate itself was used as the independent variable. Patterns in the residuals indicated that higher order variables were needed, so the square of the estimate and the natural log of the estimates were tried. Separate models were created for the three major types of estimates: crash, vehicle and person. More detailed models were also tried. For example, since almost all of the crash tables in the publication were presented for three crash severities, separate models for each severity were tested. These detailed models did not improve fit very much and were dismissed because of the added complication when computing variances. The models are presented for 1988 GES are presented in Appendix G. The R-squared values for each model are: 96 percent for the crash model, 96 percent for the vehicle model and 95 percent for the person model. No evident pattern was found in the residuals for any of the three models.

The sample design in 1989 was exactly the same as in 1988. The first and second stages of sampling did not change, only different PARs were selected. Since sampling errors should not have changed significantly, no new generalized variances were calculated. The tables in Appendix G can be used for calculating 1989 sampling errors.

The sample design in 1990 changed slightly, with the addition of a fourth stratum at the final stage of sampling. The first two stages remained the same as in 1988 and 1989. Also in 1990, computer processing of GES data by NCSA went from primarily a mainframe environment to primarily a PC-based environment. With these changes, a PC statistical package, SUDAAN (Professional Software for <u>SU</u>rvey <u>DA</u>ta <u>AN</u>alysis) was chosen to estimate sampling errors. WESVAR was not available on PC. SUDAAN employs Taylor Series Approximation (TS) to estimate sampling errors. SUDAAN was developed by the Research Triangle Institute in North Carolina[11/]. Overall, the errors calculated from TS in 1990 were very similar in size to the BRR errors calculated in 1988. New generalized variance models were created for 1990 and are presented in Appendix H. The R-squared values for these models are: 95 percent for the crash model, 98 percent for the vehicle model, and 97 percent for the person model.

As this is being written, a generalized model for proportions for the 1990 GES is being developed. The procedures to create the generalized model for the totals were followed, but an adequate model was not found. More work needs to be done to ensure that the model with best fit is found for sampling errors of proportions. If you would like more information about sampling errors of proportions, contact the NCSA.

CONFIDENCE INTERVALS

The sampling errors in Appendices G and H can be used to calculate confidence intervals about the GES estimates. The numbers in these Appendices represent an estimate of one standard error. For example, if all possible samples of PARs were selected, each of these conducted under the same conditions, then approximately two-thirds of the intervals from one standard error below the estimate to one standard error above the estimate would include the average value of all possible samples. For this particular sample, the interval would have a two-thirds chance of containing the true population value. This interval is called a 68 percent confidence interval. An interval of two standard errors above and below the estimate is a 95 percent confidence interval.

For example, from the 1989 GES, an estimated 950,000 property-damage-only crashes occurred on roads with a speed limit of 40 to 45 mph. To create a 68 percent confidence interval for this estimate, Table F1 of Appendix F should be used. If a calculator with the natural logarithmic function is not available, linear interpolation can be used. Linear interpolation may result in an estimate that is larger or smaller than the estimate from the generalized curve itself. From the standard error values for 900,000 and 1,000,000, the standard error for 950,000 is approximated at 67,700. The 68 percent confidence interval for this estimate would be 950,000 ± 67,700 or (882,300 to 1,017,700). The 95 percent confidence interval would be 950,000 ± 2*(67,700) or (814,600 to 1,085,400).

NONSAMPLING ERRORS

Nonsampling errors can be attributed to many sources other than missing data: inability of the coder to read a poor quality copy of the PAR, differences in interpreting data between coders, incorrect or conflicting information on a PAR, typing the wrong keystrokes, collection of the wrong PAR. Every effort is made during the data collection and coding process to minimize nonsampling errors. The GES Quality Control Process is implemented to control nonsampling errors. The major components of this process are described below. Copies of the latest GES data edits and "General Estimates System Quality Control Plan for Data Entry, March 1988." [12/]

Prior to Data Entry:

> Sampling Process Quality Control: Visits are made to each PSU at least once a year by trained NHTSA or contractor personnel to check that all PARs are being listed, that PARs are being stratified properly, and that the correct sampling intervals are being used. In addition, once the PARs are sent to the data coding contractor, the intervals and carryover numbers used in the stratified sample are checked. The sampled PAR number is compared with the actual PAR that was sent.

During Data Entry:

> Range Checks: These checks examine the data as they are entered to determine whether the entry is the correct

type (numeric or character) and whether it is in the acceptable range for that variable. For example, time of day in military time cannot be 36:42. If any data fail a range check, the data coder can not continue to enter data until the correct code is entered.

Intraform Checks: These checks examine the logical links between the variables for one of the five data entry sets-- accident, vehicle, driver, occupant, and nonmotorist-- after all data for that set have been entered. For example, an intraform check would fail if the crash variable *time of day* is coded "midnight" and the crash variable *light condition* is coded "daylight." When data fail an intraform check, the data in question are reviewed and changes must be made before entering any data for the next data entry set.

Interform Checks: These checks examine the logical links between data in two or more data entry sets after all data for a case have been entered. For example, an interform check would fail if the crash variable *first harmful event* is coded "collision with a pedestrian" then at least on person must have a *person type* of "pedestrian" on the person level. If any data fail these checks, those data must be reviewed and changed before the case can be completed.

Intercoder Reliability: This process attempts to ensure that all coders interpret information in the same manner. Intercoder reliability is accomplished through training and recoding by the data coding supervisor. Recoding is performed on a periodic basis depending on the level of errors for that coder in the past. Any errors found by the coding supervisor are discussed with the coder and changes are made.

During File Creation:

Records Verification: As data are uploaded each day and when the SAS file is created, several records verification programs are run to ensure that no records are lost or added. For example, if the case has two vehicles the uploaded file is checked to make sure there are two vehicle records for that case.

After File Creation:

Data Consistency Checks - On-site: The edits described above are re-run on the entire file and any data errors that may have slipped through earlier checks are corrected. In addition, ad hoc programs such as frequency counts are run to look for data inconsistencies. Sometimes new edit checks are developed during the year and this step ensures that all edits have been applied to all data.

Data Consistency Checks - Off-site: NCSA conducts its own quality review after receiving the data file. SAS programs are run, including some of the annual report tabulations, to identify large inconsistencies from previous year's data.

VIII. CLOSING NOTES

The procedures discussed in this Technical Note are subject to change. Every second year GES variables are revised to better serve the data users. During the collection year, new edits are added to the data entry system as variable relationships are better understood. New imputation methods continue to be tested. Each year the sampling errors are recalculated. This Note will be amended as significant changes occur to GES. Please contact NCSA at the address given in the Introduction for any other information about GES.

VIV. REFERENCES

1/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *National Accident Sampling System 1990 Crashworthiness Data System Data Collection, Coding, and Editing Manual*, (January, 1990), pp. 5-39.

2/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *National Accident Sampling System Sample Design, Phases 2 and 3, Volume 1: Final Technical Report*, (November, 1979), DOT HS-805-274.

3/ Keyfitz, Nathan, "Sampling with Probabilities Proportional to Size," Journal of the American Statistical Association, 46 (March, 1951), pp. 105-109.

4/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *National Accident Sampling System General Estimates System Researcher Manual*, (January, 1990), pp. 4-39.

5/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *General Estimates System 1988: A Review of Information on Police-Reported Traffic Crashes in the United States*, (August, 1990), DOT HS-807-607.

6/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *General Estimates System 1989: A Review of Information on Police-Reported Traffic Crashes in the United States*, (December, 1990), DOT HS-807-665.

7/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *General Estimates System 1990: A Review of Information on Police-Reported Traffic Crashes in the United States*, (November, 1991), DOT HS-807-781.

8/ See, for example, Little, Roderick J.A. and Rubin, Donald B., Statistical Analysis with Missing Data, (1987), Wiley, pp. 62-67.

9/ Westat, Inc., *The Wesvar Procedure, B Version*, (May, 1989), Rockville, Maryland.

10/ See, for example, Wolter, Kirk M., Introduction to Variance Estimation, (1985), Springer-Verlag, pp. 110-152.

11/ Research Triangle Institute, *SUDAAN Professional Software for SUrvey DAta ANalysis*, (1989), Research Triangle Park, North Carolina.

12/ National Highway Traffic Safety Administration, National Center for Statistics and Analysis, *General Estimates System Quality Control Plan for Data Entry*, (March, 1988).

APPENDICES

APPENDICES

APPENDIX A

ABBREVIATIONS

BRR — **Balanced Repeated Replication** - commonly used variance estimation method used in surveys with complex sample designs. Used to estimate variances for the 1988 and 1989 General Estimates System.

CDS — **Crashworthiness Data System** - with the General Estimates System, replaced the "old" National Accident Sampling System in 1988. Collects detailed information on vehicles and occupants in motor vehicle crashes involving a towed passenger car, light truck, light van or utility vehicle.

GES — **General Estimates System** - with the Crashworthiness Data System, replaced the National Accident Sampling System in 1988. Collects police-reported crash data on all types of motor vehicle crashes.

NASS — **National Accident Sampling System** - a survey of motor vehicle traffic crashes in the United States that began in 1979 as one system. Divided into two systems in 1988: NASS CDS and NASS GES.

NCSA — **National Center for Statistics and Analysis** - Office within the National Highway Traffic Safety Administration, whose responsibilities include the operation of the NASS CDS and NASS GES.

NHTSA — **National Highway Traffic Safety Administration** - Federal agency whose mission is to reduce the human and economic costs of our motor vehicle and highway traffic system.

PAR — **Police Accident Report** - report that is filled out by a police officer at the scene of a motor vehicle crash and is the sole source of data for the GES.

PDO — **Property Damage Only** - motor vehicle crash that results in no injuries to anyone involved in the crash.

PJ — **Police Jurisdiction** - sites where PARs for GES sample are collected.

PSU — **Primary Sampling Unit** - geographic areas that are selected for first stage of GES sampling.

TS — **Taylor Series Approximation** - variance estimation method used for surveys with complex sample designs. Used to calculate variances for the 1990 GES.

APPENDIX B

1990 GENERAL ESTIMATES SYSTEM BODYTYPES

Automobiles

01 = Convertible (excludes sun-roof, t-bar)
02 = 2-door sedan, hardtop, coupe
03 = 3-door/2-door hatchback
04 = 4-door sedan, hardtop
05 = 5-door/4-door hatchback
06 = Station wagon (excluding van and truck based)
07 = Hatchback, number of doors unknown
08 = Other automobile type
09 = Unknown automobile type

Automobile Derivatives

10 = Auto based pickup
11 = Ambulance
12 = Hearse
13 = Limousine

Utility Vehicles

14 = Utility vehicles

Van Based Light Trucks (< 10,000 lbs GVWR)

20 = Minivan
21 = Standard Van
22 = Step-van or Walk-in van
28 = Other van type
29 = Unknown van type

Light Conventional Trucks (Pickup style cab, < 10,000 lbs GVWR)

30 = Compact pickup
31 = Standard pickup
32 = Pickup with slide-in camper
39 = Unknown (pickup style) light conventional truck

Other Light Trucks (< 10,000 lbs GVWR)

40= Cab chassis based (includes rescue vehicle, light stake, dump, and tow truck)
41= Truck based panel
42= Light truck based motorhome (chassis mounted)
48= Unknown light truck type (pickup, van or other)
49= Unknown light vehicle type (automobile, van, or light truck)

Buses

50= School bus type (designed to carry students, not cross country or transit)
58= Other bus (e.g., transit, intercity, bus based motorhome)
59= Unknown bus type

Medium/Heavy Trucks (>10,000 lbs GVWR)

60= Single unit straight truck
63= Medium/heavy truck based motorhome
65= Truck-tractor (Cab Only, or with any number of trailing units; any weight)
68= Unknown medium/heavy truck type
69= Unknown truck type (light/medium/heavy)

Motored Cycles (Does not include all terrain vehicles/cycles)

70= Motorcycle
71= Moped (motorized bicycle)
72= Three wheeled motorcycle or moped
78= Other motored cycle type (minibike, motorscooter)
79= Unknown motored cycle type

Other Vehicles

80= ATV (all terrain vehicle) and ATC (all terrain cycle)
81= Snowmobile
82= Farm equipment other than trucks
83= Construction equipment other than trucks (includes graders)
88= Other type vehicle (includes go-cart, fork lift, city street sweeper)
89= Unknown other vehicle
99= Unknown body type

APPENDIX C

GENERAL ESTIMATES SYSTEM

PRIMARY SAMPLING UNITS

ALABAMA

1 Bibb and Tuscaloosa Counties
2 Chilton, Coosa, St. Clair, Shelby and Talladega Counties

ARIZONA

3 Gila, Graham, and Greenlee Counties
4 Pima County
5 Yuma and La Paz Counties

CALIFORNIA

6 Contra Costa County
7 Los Angeles
8 Los Angeles County (except cities of Los Angeles and Long Beach)
9 Ventura County

COLORADO

10 Boulder County
11 Gilpin and Jefferson Counties

FLORIDA

12 Dade County (except city of Miami)
13 Fort Lauderdale and Hollywood
14 Sarasota County

IOWA

15 Bremer, Butler, Chickasaw, Floyd, Grundy, Howard, and Mitchell Counties

ILLINOIS

16 Chicago
17 Cook County (except city of Chicago)

INDIANA

18 Lake County

KENTUCKY

19 Harlan and Letcher Counties

MASSACHUSETTS

20 Hampshire County
21 Middlesex County

MARYLAND

22 Charles and Prince Georges Counties

MICHIGAN

23 Detroit
24 Genesee County
25 Muskegon County
26 Oakland County
27 Wastenaw County
28 Wayne County (except city of Detroit)

MISSOURI

29 St. Louis County

NORTH CAROLINA

30 Cleveland and Rutherford Counties
31 Wake County

NEBRASKA

32 Douglas County

NEW JERSEY

33 Ocean County

NEW MEXICO

34 Bernalillo and Sandoval Counties

NEW YORK

35 Erie County (except city of Buffalo
36 Kings County
37 New York County
38 Queens County
39 Suffolk County
40 Ulster County
41 Schenectady County (except city of Schenectady)

OHIO

42 Logan and Shelby Counties
43 Columbus
44 Preble and Warren Counties

OKLAHOMA

45 Oklahoma City

PENNSYLVANIA

46 Delaware County
47 Philadephia
48 Montgomery County
49 Allegheny County (except city of Pittsburgh

TENNESSEE

50 Knox County
51 Shelby and Tipton Counties (except city of Memphis)

TEXAS

52 Brazoria County
53 Dallas
54 Dallas County (except of Dallas)
55 Houston
56 Midland County

VIRGINIA

57 Henrico County and Richmond

WASHINGTON

58 King County (except of Seattle)
59 Seattle

WISCONSIN

60 Waukesha County

APPENDIX D

1990 GENERAL ESTIMATES SYSTEM

FIRST AND SECOND STAGE SAMPLING WEIGHTS

PSU	PJ	PSU WGT	PJ WGT	PSU	PJ	PSU WGT	PJ WGT
1	1	39.95	1.00	4	6	13.77	3.77
1	2	39.95	1.00	4	7	13.77	7.43
1	3	39.95	1.00	4	8	13.77	12.49
1	4	39.95	1.00	5	1	5.04	1.00
1	5	39.95	1.00	5	2	5.04	1.78
1	6	39.95	1.00	5	3	5.04	2.66
1	7	39.95	1.00	5	4	5.04	2.97
1	8	39.95	1.00	5	5	5.04	4.29
2	1	24.52	1.00	5	6	5.04	3.70
2	2	24.52	1.00	5	7	5.04	4.66
2	3	24.52	1.00	5	8	5.04	5.50
2	4	24.52	2.03	5	9	5.04	10.50
2	5	24.52	2.16	6	1	2.39	1.96
2	6	24.52	2.34	6	2	2.39	1.98
2	7	24.52	2.77	6	3	2.39	1.97
2	8	24.52	5.24	6	4	2.39	1.91
3	1	1.27	2.00	6	5	2.39	2.71
3	2	1.27	4.62	6	6	2.39	2.84
3	3	1.27	3.90	6	7	2.39	2.83
3	4	1.27	4.02	6	8	2.39	3.95
3	5	1.27	5.72	6	9	2.39	7.20
3	6	1.27	5.92	7	1	8.12	1.00
3	7	1.27	5.76	7	2	8.12	1.00
3	8	1.27	5.42	7	3	8.12	2.25
3	9	1.27	8.06	7	4	8.12	2.16
4	1	13.77	1.00	7	5	8.12	2.58
4	2	13.77	1.00	7	6	8.12	4.69
4	3	13.77	1.00	7	7	8.12	7.87
4	4	13.77	2.37	7	8	8.12	20.04
4	5	13.77	2.93	8	1	4.80	1.00

PSU	PJ	PSU WGT	PJ WGT
8	2	4.80	1.76
8	3	4.80	2.92
8	4	4.80	2.98
8	5	4.80	4.99
8	6	4.80	8.03
8	7	4.80	9.37
8	8	4.80	12.89
8	9	4.80	23.81
8	10	4.80	8.60
8	11	4.80	8.60
8	12	4.80	8.60
8	13	4.80	12.92
8	14	4.80	12.92
9	1	3.96	1.00
9	2	3.96	1.00
9	3	3.96	1.00
9	4	3.96	1.00
9	5	3.96	1.00
9	6	3.96	1.00
9	7	3.96	1.78
9	8	3.96	2.34
9	9	3.96	6.45
10	1	2.50	1.00
10	2	2.50	1.80
10	3	2.50	3.24
10	4	2.50	2.94
10	5	2.50	3.80
10	6	2.50	5.29
10	7	2.50	5.29
10	8	2.50	8.76
10	9	2.50	58.48
11	1	20.07	1.00
11	2	20.07	1.00
11	3	20.07	1.00
11	4	20.07	1.00
11	5	20.07	1.00
11	6	20.07	3.14
11	7	20.07	4.08
12	1	6.03	1.00
12	2	6.03	1.00
12	3	6.03	1.00
12	4	6.03	1.94
12	5	6.03	2.58
12	6	6.03	4.87
12	7	6.03	12.39
13	1	33.68	1.00
13	2	33.68	1.00
13	3	33.68	1.00
13	4	33.68	1.00
13	5	33.68	1.00
13	6	33.68	2.42
14	1	4.22	1.00
14	2	4.22	1.00
14	3	4.22	1.00
14	4	4.22	1.00
14	5	4.22	2.04
14	6	4.22	1.78
14	7	4.22	8.96
14	8	4.22	8.96
14	9	4.22	5.61
14	10	4.22	5.61
15	1	37.75	1.00
15	2	37.75	1.00
15	3	37.75	1.87
15	4	37.75	1.00
15	5	37.75	2.67
15	6	37.75	3.06
16	1	1.88	10.00
16	2	1.88	1.00
16	3	1.88	2.11
16	4	1.88	3.22
16	5	1.88	10.08
17	1	1.47	4.44
17	2	1.47	3.88
17	3	1.47	6.10
17	4	1.47	6.36
17	5	1.47	8.48
17	6	1.47	1.91
18	1	1.86	2.00
18	2	1.86	4.06

PSU	PJ	PSU WGT	PJ WGT	PSU	PJ	PSU WGT	PJ WGT
18	3	1.86	5.42	26	7	2.32	6.25
18	4	1.86	2.95	26	8	2.32	19.26
18	5	1.86	5.35	27	1	5.35	1.00
18	6	1.86	4.56	27	2	5.35	1.00
19	1	5.60	1.00	27	3	5.35	1.00
19	2	5.60	2.31	27	4	5.35	1.00
19	3	5.60	2.89	28	1	1.67	2.00
19	4	5.60	4.59	28	2	1.67	2.00
19	5	5.60	5.89	28	3	1.67	2.00
19	6	5.60	19.73	28	4	1.67	1.00
20	1	10.21	1.00	28	5	1.67	1.00
20	2	10.21	1.00	28	6	1.67	3.02
20	3	10.21	1.00	28	7	1.67	8.54
20	4	10.21	1.00	29	1	16.38	1.00
21	1	110.12	1.00	29	2	16.38	1.00
21	2	110.12	1.00	29	3	16.38	1.00
21	3	110.12	1.00	29	4	16.38	1.00
21	4	110.12	1.00	29	5	16.38	4.02
21	5	110.12	1.00	29	6	16.38	19.01
22	1	17.50	1.00	30	1	46.64	1.00
22	2	17.50	1.00	30	2	46.64	1.00
22	3	17.50	2.08	30	3	46.64	1.00
22	4	17.50	1.62	30	4	46.64	1.00
22	5	17.50	3.77	30	5	46.64	1.00
23	1	4.32	3.33	30	6	46.64	9.89
23	2	4.32	1.00	31	1	9.88	1.00
23	3	4.32	1.00	31	2	9.88	1.00
24	1	92.62	1.00	31	3	9.88	1.00
24	2	92.62	1.00	32	1	29.83	1.00
24	3	92.62	1.00	32	2	29.83	1.00
24	4	92.62	1.00	32	3	29.83	1.00
24	5	92.62	1.00	32	4	29.83	1.95
25	1	1.90	10.00	32	5	29.83	1.94
25	2	1.90	1.00	32	6	29.83	1.63
26	1	2.32	10.00	33	1	30.37	1.00
26	2	2.32	10.08	33	2	30.37	1.00
26	3	2.32	4.58	33	3	30.37	1.00
26	4	2.32	3.60	33	4	30.37	1.73
26	5	2.32	3.59	33	5	30.37	2.02
26	6	2.32	3.38	33	6	30.37	3.67

PSU	PJ	PSU WGT	PJ WGT
33	7	30.37	11.38
34	1	54.65	1.00
34	2	54.65	1.00
34	3	54.65	1.00
34	4	54.65	1.00
34	5	54.65	1.00
34	6	54.65	1.00
34	7	54.65	1.00
34	8	54.65	1.00
34	9	54.65	1.00
35	1	2.88	2.00
35	2	2.88	1.00
36	1	5.25	1.00
36	2	5.25	1.00
36	3	5.25	1.00
36	4	5.25	1.95
36	5	5.25	2.41
36	6	5.25	3.11
36	7	5.25	7.28
37	1	47.72	1.00
37	2	47.72	1.00
37	3	47.72	1.00
38	1	24.31	1.00
38	2	24.31	1.00
38	3	24.31	1.00
38	4	24.31	1.00
38	5	24.31	1.00
38	6	24.31	1.00
39	1	19.32	1.00
39	2	19.32	1.00
39	3	19.32	1.00
39	4	19.32	1.81
39	5	19.32	2.06
39	6	19.32	2.02
39	7	19.32	8.12
41	1	1.64	10.00
41	2	6.06	2.00
41	3	6.06	1.00
42	1	13.45	1.00
42	2	13.45	1.00
42	3	13.45	1.00
42	4	13.45	1.00
42	5	13.45	1.64
42	6	13.45	3.48
42	7	13.45	11.51
43	1	1.00	6.00
43	2	1.00	6.00
43	3	1.00	6.00
43	4	1.00	6.00
43	5	1.00	6.00
43	6	1.00	6.00
43	7	1.00	2.00
43	8	1.00	2.00
44	1	5.54	1.00
44	2	5.54	1.00
44	3	5.54	1.00
44	4	5.54	1.73
44	5	5.54	2.54
44	6	5.54	2.73
44	7	5.54	11.29
45	1	5.53	1.00
45	2	5.53	1.00
45	3	5.53	1.00
45	4	5.53	1.00
45	5	5.53	1.00
45	6	5.53	1.00
45	7	5.53	1.00
45	8	5.53	1.00
46	1	6.59	1.00
46	2	6.59	1.00
46	3	6.59	1.00
46	4	6.59	1.00
46	5	6.59	3.14
47	1	108.19	1.00
47	2	108.19	1.00
47	3	108.19	1.00
47	4	108.19	1.00
47	5	108.19	2.47
47	6	108.19	3.61
48	1	3.92	2.00

PSU	PJ	PSU WGT	PJ WGT	PSU	PJ	PSU WGT	PJ WGT
48	2	3.92	1.00	54	8	1.09	12.05
48	3	3.92	1.00	54	9	1.09	11.52
48	4	3.92	1.00	54	10	1.09	64.64
48	5	3.92	1.00	54	11	1.09	64.64
48	6	3.92	1.00	54	12	1.09	64.64
48	7	3.92	1.00	55	1	3.86	1.00
49	1	58.33	1.00	55	2	3.86	3.65
49	2	58.33	1.00	55	3	3.86	7.39
49	3	58.33	1.00	55	4	3.86	8.48
49	4	58.33	1.00	55	5	3.86	2.00
49	5	58.33	4.24	55	6	3.86	2.74
50	1	1.52	5.71	55	7	3.86	4.20
50	2	1.52	2.00	55	8	3.86	4.49
50	3	1.52	3.40	55	9	3.86	5.01
50	4	1.52	1.24	55	10	3.86	5.21
51	1	4.89	1.00	55	11	3.86	5.82
51	2	4.89	1.00	55	12	3.86	6.79
51	3	4.89	1.00	56	1	89.40	1.00
51	4	4.89	1.00	56	2	89.40	1.00
51	5	4.89	2.10	56	3	89.40	1.00
51	6	4.89	1.95	56	4	89.40	1.00
51	7	4.89	2.17	56	5	89.40	1.00
51	8	4.89	7.39	56	6	89.40	1.00
52	1	3.19	2.00	56	7	89.40	1.00
52	2	3.19	2.00	56	8	89.40	1.00
52	3	3.19	1.00	56	9	89.40	2.95
52	4	3.19	1.00	56	10	89.40	4.17
52	5	3.19	1.69	56	11	89.40	5.95
52	6	3.19	3.06	57	1	13.00	1.00
52	7	3.19	4.25	57	2	13.00	1.00
52	8	3.19	44.87	57	3	13.00	1.00
53	1	6.75	1.00	57	4	13.00	2.95
53	2	6.75	1.00	57	5	13.00	1.42
54	1	1.09	5.33	58	1	4.10	2.00
54	2	1.09	9.16	58	2	4.10	1.00
54	3	1.09	9.21	58	3	4.10	1.00
54	4	1.09	9.21	58	4	4.10	2.29
54	5	1.09	9.21	58	5	4.10	15.17
54	6	1.09	7.16	59	1	1.00	22.57
54	7	1.09	9.57	59	2	1.00	33.83

PSU	PJ	PSU WGT	PJ WGT
59	3	1.00	9.02
59	4	1.00	14.24
59	5	1.00	10.68
59	6	1.00	7.34
59	7	1.00	11.32
59	8	1.00	14.28
59	9	1.00	20.35
59	10	1.00	11.69
60	1	6.48	3.33
60	2	6.48	3.33
60	3	6.48	3.98
60	4	6.48	1.00
60	5	6.48	4.12

APPENDIX E

1990 PAR STRATIFICATION RECORD

NASS
GENERAL ESTIMATES SYSTEM

PAR STRATIFICATION RECORD — Page 1 of 2

PSU: 91 | JURISDICTION: MSP #29 | LISTED BY: DOT | RECONTACT DATE: 12 / 31 / 90

PAR INFORMATION			(1) NO TRUCKS	(2) MEDIUM OR HEAVY TRUCKS		(3) OTHER INJURY	(4) OTHER
			CDS Stratum Interval: 3	CDS and X Strata Interval: 1		Y Stratum Interval: 2	Z Stratum Interval: 4
			Carryover: 4	Carryover: 1		Carryover: 2	Carryover: 4
Date	Time	Number	Line Number	Line Number	Stratum CDS or X	Line Number	Line Number
11 — 01 — 90	2145	15229	5				
11 — 02 — 90	0720	15245	6				
11 — 03 — 90	1050	15261					1
12 — 05 — 90	1600	15267				1	
12 — 07 — 90	1305	15275					2
12 — 10 — 90	0845	15312					3
12 — 12 — 90	0910	15275					4
12 — 15 — 90	1031	15312		1	CDS		
12 — 15 — 90	1225	15317	7				
12 — 17 — 90	2225	15322					1
12 — 18 — 90	2030	15323					2
12 — 20 — 90	2105	15325	8				
12 — 22 — 90	2135	15330		1	X		
12 — 23 — 90	2030	15331	1				
12 — 24 — 90	1810	15342					3
12 — 26 — 90	0500	15343					4
12 — 27 — 90	1123	15344	2				
12 — 29 — 90	1525	15349					1
12 — 30 — 90	1730	15360				2	
12 — 30 — 90	1920	15361					2

	CDS (No Trucks)	CDS (Trucks)	X	Y	Z
Page Totals:	6	1	1	2	10
Day Totals:					

APPENDIX F

GENERAL ESTIMATES SYSTEM VARIABLES AND PERCENTAGE UNKNOWN

	GES CRASH VARIABLES	1988 UNKNOWN %	1989 UNKNOWN %	1990 UNKNOWN %
1	ALCOHOL INVOLVED	4.8	4.0	3.1
2	ATMOSPHERIC CONDITIONS	2.7	2.4	2.0
3	DAY	0.1	0.0	0.0
4	FIRST HARMFUL EVENT	0.3	0.3	0.4
5	HOUR	1.5	1.2	0.9
6	INTERSTATE HIGHWAY	0.7	0.1	0.1
7	LAND USE	2.8	4.5	4.7
8	LIGHT CONDITION	3.1	2.5	2.7
9	MANNER OF COLLISION	1.7	0.9	0.8
10	MAXIMUM INJURY SEVERITY	3.6	3.7	2.5
11	MINUTE	1.4	1.2	0.9
12	MONTH	0.0	0.0	0.0
13	NUMBER INJURED	*	*	*
14	NUMBER OF NON-MOTORISTS	*	*	*
15	NUMBER OF TRAVEL LANES	27.9	26.6	28.8
16	NUMBER OF VEHICLES	*	*	*
17	PEDESTRIAN/BICYCLIST ACCIDENT TYPE	0.0	0.0	0.0
18	PERCENTAGE RURAL	*	*	*
19	POLICE JURISDICTION	*	*	*
20	PRIMARY SAMPLING UNIT	*	*	*
21	CENSUS REGION	**	**	*
22	RELATION TO JUNCTION	4.5	1.1	1.2

23	RELATION TO ROADWAY	1.0	0.4	0.4
24	ROADWAY ALIGNMENT	5.2	3.5	3.4
25	ROADWAY PROFILE	32.5	31.0	31.2
26	ROADWAY SURFACE CONDITION	1.9	1.6	1.8
27	SCHOOL BUS RELATED	*	*	*
28	SPEED LIMIT	35.0	34.3	32.6
29	STATE	*	*	*
30	TRAFFIC CONTROL DEVICE	3.4	2.7	2.6
31	TRAFFICWAY FLOW	26.7	31.1	31.6
32	WEIGHT	*	*	*
33	YEAR	*	*	*

* By definition of the variable, no unknowns could be coded.

	GES VEHICLE VARIABLES	1988 UNKNOWN %	1989 UNKNOWN %	1990 UNKNOWN %
1	ACCIDENT TYPE	2.6	1.4	1.2
2	BODY TYPE	5.4	4.8	4.2
3	DAMAGE AREAS	*	*	4.9
4	DAMAGE SEVERITY	24.2	27.2	27.0
5	DRIVER DISTRACTED BY	*	*	0.3
6	DRIVER DRINKING IN VEHICLE	6.1	5.8	6.7
7	DRIVER MANUEVERED TO AVOID	1.9	0.7	0.3
8	DRIVER PRESENCE	0.3	0.1	0.0
9	DRIVERS VISION OBSCURED BY	1.9	1.1	0.6
10	EMERGENCY USE	0.9	0.1	0.0
11	FIRE OCCURRENCE	0.2	0.2	0.2
12	HIT AND RUN	0.2	0.1	0.1
13	INITIAL POINT OF IMPACT	12.2	6.7	5.0
14	JACKKNIFE	0.2	0.1	0.1
15	MANNER OF LEAVING SCENE	11.3	9.7	5.6
16	MAXIMUM INJURY SEVERITY	4.7	5.2	5.8
17	MODEL YEAR	6.0	5.6	5.8
18	MOST HARMFUL EVENT	2.0	3.3	3.6
19	NUMBER INJURED IN VEHICLE	0.1	0.5	0.4
20	NUMBER OF OCCUPANTS	**	**	**
21	ROLLOVER	2.1	2.2	2.3
22	SPECIAL USE	1.0	0.1	0.1
23	TRAVEL SPEED	64.1	61.3	60.8
24	VEHICLE DEFECTS	5.0	2.3	1.3
25	VEHICLE MAKE	5.4	5.3	5.2
26	VEHICLE MANEUVER	2.6	2.3	2.1

27	VEHICLE MODEL	55.0	53.8	52.8
28	VEHICLE ROLE	0.9	1.3	1.1
29	VEHICLE TRAILING	1.1	0.2	0.1
30	VIN LISTED	33.2	33.5	31.7
31	VIOLATIONS CHARGED	2.2	1.1	0.4

* No comparable 1988 or 1989 variables

** By definition of the variable, no unknowns could be coded.

	GES PERSON VARIABLES	1988 UNKNOWN %	1989 UNKNOWN %	1990 UNKNOWN %
1	AGE	13.9	13.0	13.0
2	EJECTION	0.6	0.2	0.1
3	INJURY SEVERITY	4.2	4.7	4.7
4	NON-MOTORIST LOCATION	0.1	0.0	0.0
5	NON-MOTORIST'S ACTION	0.7	0.1	0.0
6	NON-MOTORIST'S SAFETY EQUIPMENT USE	*	*	0.4
7	PERSON TYPE	0.4	0.1	0.1
8	PERSONS PHYSICAL IMPAIRMENT	*	*	3.5
9	POLICE REPORTED ALCOHOL INVOLVEMENT	4.5	4.1	4.3
10	POLICE REPORTED DRUG INVOLVEMENT	*	*	4.4
11	RESTRAINT SYSTEM USE	20.4	19.1	18.6
12	RESTRAINT TYPE	*	*	64.3
13	SEATING POSITION	4.3	2.9	2.9
14	SEX	11.4	10.7	10.6
15	TAKEN TO HOSPITAL OR TREATMENT FACILITY	2.3	1.7	1.6

* No comparable 1988 or 1989 variables existed.

APPENDIX G

GENERALIZED ESTIMATED SAMPLING ERRORS FOR 1988 AND 1989 GES

TABLE G1:

1988 AND 1989 CRASH ESTIMATES AND STANDARD ERRORS

Estimate (x)	One Standard Error (SE)*	Estimate (x)	One Standard Error (SE)*
1,000	600	600,000	44,800
5,000	1,400	700,000	51,300
10,000	2,100	800,000	57,900
20,000	3,200	900,000	64,400
30,000	4,200	1,000,000	71,000
40,000	5,200	1,500,000	103,700
50,000	6,100	2,000,000	136,500
60,000	6,900	2,500,000	169,600
70,000	7,800	3,000,000	203,100
80,000	8,600	3,500,000	236,900
90,000	9,400	4,000,000	271,000
100,000	10,200	4,500,000	305,400
200,000	17,600	5,000,000	340,200
300,000	24,600	5,500,000	375,400
400,000	31,400	6,000,000	410,800
500,000	38,100	7,000,000	482,600

$$*SE=e^{\frac{a}{2}+\frac{b}{2}[\ln(x)]^2}, where$$
$$a=9.62798$$
$$b=0.06660$$

TABLE G2:

1988 AND 1989 VEHICLE ESTIMATES AND STANDARD ERRORS

Estimate (x)	One Standard Error (SE)*	Estimate (x)	One Standard Error (SE)*
1,000	500	600,000	43,400
5,000	1,200	700,000	50,000
10,000	1,800	800,000	56,600
20,000	2,900	900,000	63,200
30,000	3,800	1,000,000	69,900
40,000	4,700	2,000,000	137,400
50,000	5,500	3,000,000	207,300
60,000	6,300	4,000,000	279,300
70,000	7,100	5,000,000	353,400
80,000	7,900	6,000,000	429,500
90,000	8,600	7,000,000	507,300
100,000	9,400	8,000,000	586,800
200,000	16,500	9,000,000	667,900
300,000	23,400	10,000,000	750,500
400,000	30,100	11,000,000	834,500
500,000	36,700	12,000,000	919,900

$$*SE=e^{\frac{a}{2}+\frac{b}{2}[\ln(x)]^2}, where$$
$$a=9.16179$$
$$b=0.06888$$

TABLE G3:

1988 AND 1989 PERSON ESTIMATES AND STANDARD ERRORS

Estimate (x)	One Standard Error (SE)*	Estimate (x)	One Standard Error (SE)*
1,000	500	600,000	45,400
5,000	1,200	700,000	52,500
10,000	1,800	800,000	59,500
20,000	2,900	900,000	66,600
30,000	3,800	1,000,000	73,800
40,000	4,700	2,000,000	146,800
50,000	5,600	3,000,000	223,000
60,000	6,400	4,000,000	302,200
70,000	7,200	5,000,000	384,000
80,000	8,000	6,000,000	468,200
90,000	8,800	7,000,000	554,700
100,000	9,500	8,000,000	643,300
200,000	17,000	9,000,000	733,900
300,000	24,200	10,000,000	826,300
400,000	31,300	11,000,000	920,600
500,000	38,300	12,000,000	1,016,600

$$*SE=e^{\frac{a}{2}+\frac{b}{2}[\ln(x)]^2}, where$$
$$a=9.03591$$
$$b=0.07011$$

APPENDIX H

GENERALIZED ESTIMATED SAMPLING ERRORS FOR 1990 GES

TABLE H1:

1990 CRASH ESTIMATES AND STANDARD ERRORS

Estimate (x)	One Standard Error (SE)*	Estimate (x)	One Standard Error (SE)*
1,000	700	600,000	40,000
5,000	1,400	700,000	45,700
10,000	2,100	800,000	51,200
20,000	3,300	900,000	56,700
30,000	4,200	1,000,000	62,200
40,000	5,100	2,000,000	116,200
50,000	5,900	3,000,000	169,800
60,000	6,800	4,000,000	223,700
70,000	7,500	5,000,000	278,000
80,000	8,300	6,000,000	332,800
90,000	9,000	7,000,000	388,100
100,000	9,700	8,000,000	444,000
200,000	16,400	9,000,000	500,400
300,000	22,600	10,000,000	557,300
400,000	28,600	11,000,000	614,700
500,000	34,400	12,000,000	672,500

$$*SE=e^{\frac{a}{2}+\frac{b}{2}[\ln(x)]^2}, where$$
$$a=9.93401$$
$$b=0.06362$$

TABLE H2:

1990 VEHICLE ESTIMATES AND STANDARD ERRORS

Estimate (x)	One Standard Error (SE)*	Estimate (x)	One Standard Error (SE)*
1,000	400	600,000	39,900
5,000	1,000	700,000	46,100
10,000	1,600	800,000	52,200
20,000	2,500	900,000	58,400
30,000	3,400	1,000,000	64,700
40,000	4,200	2,000,000	128,300
50,000	4,900	3,000,000	194,500
60,000	5,700	4,000,000	263,100
70,000	6,400	5,000,000	334,000
80,000	7,100	6,000,000	406,900
90,000	7,800	7,000,000	481,600
100,000	8,500	8,000,000	558,200
200,000	15,000	9,000,000	636,400
300,000	21,300	10,000,000	716,100
400,000	27,500	11,000,000	797,400
500,000	33,700	12,000,000	880,100

$$*SE=e^{\frac{a}{2}+\frac{b}{2}[\ln(x)]^2}, where$$
$$a=8.83524$$
$$b=0.06977$$

TABLE H3:

1990 PERSON ESTIMATES AND STANDARD ERRORS

Estimate (x)	One Standard Error (SE)*	Estimates	One Standard Error (SE)*
1,000	400	600,000	34,800
5,000	1,000	700,000	40,100
10,000	1,500	800,000	45,300
20,000	2,400	900,000	50,600
30,000	3,100	1,000,000	55,800
40,000	3,900	2,000,000	108,800
50,000	4,500	3,000,000	163,200
60,000	5,200	4,000,000	219,100
70,000	5,800	5,000,000	276,400
80,000	6,500	6,000,000	335,000
90,000	7,100	7,000,000	394,900
100,000	7,700	8,000,000	455,900
200,000	13,400	9,000,000	518,100
300,000	18,900	10,000,000	581,300
400,000	24,300	11,000,000	645,500
500,000	29,600	12,000,000	710,600

$$*SE=e^{\frac{a}{2}+\frac{b}{2}[\ln(x)]^2}, where$$
$$a=8.88000$$
$$b=0.068000$$